HYDROLOGICAL PROCESSES
Hydrol. Process. **24**, 1048–1062 (2010)
Published online 13 January 2010 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.7546

Comparison of spectral and time domain calibration methods for precipitation-discharge processes

Jan J. Quets, Gabriëlle J. M. De Lannoy* and Valentijn R. N. Pauwels
Laboratory of Hydrology and Water Management, Ghent University, Coupure links 653, B-9000 Ghent, Belgium

Abstract:

Hydrological model parameter estimation is an important aspect in hydrologic modelling. Usually, parameters are estimated through an objective function minimization, quantifying the mismatch between the model results and the observations. The objective function choice has a large impact on the sensitivity analysis and calibration outcomes. In this study, it is assessed whether spectral objective functions can compete with an objective function in the time domain for optimization of the Soil and Water Assessment Tool (SWAT). Three empirical spectral objective functions were applied, based on matching (i) Fourier amplitude spectra, (ii) periodograms and (iii) Fourier series of simulated and observed discharge time series. It is shown that most sensitive parameters and their optimal values are distinct for different objective functions. The best results were found through calibration with an objective function based on the square difference between the simulated and observed discharge Fourier series coefficients. The potential strengths and weaknesses of using a spectral objective function as compared to utilising a time domain objective function are discussed. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS calibration; discharge; Fourier analysis; objective function; sensitivity analysis; spectral calibration; equifinality

Received 5 November 2008; Accepted 23 October 2009

INTRODUCTION

Along with the development of physically based models in hydrology, the estimation of model parameters became an important issue. Ideally, measured values should be assigned to the model parameters. However, measured parameters can deviate significantly from the effective model parameters (Mertens *et al*., 2005), because the measurement scale (mostly point scale) deviates from the model application scale (e.g. hydrological response units) (Beven, 1995). For a complex physically based model, such as the Soil and Water Assessment Tool (SWAT) (Arnold *et al*., 1998), which is characterized by a multiplicity of spatially distributed parameters, it is nearly impossible to measure all model parameters. Different model calibration methods are available, ranging from simple manual calibration (trial and error) (Boyle *et al*. 2000) to more complex automatic calibrations, which attempt to minimize a predefined objective function using a numeric algorithm. A large number of methods have been developed for this purpose, for example, the parameter estimation (PEST) method (Doherty, 2001), the shuffled complex evolution (SCE-UA) algorithm (Duan *et al*., 1992, 1994; Yapo *et al*., 1998), genetic algorithms (Reed *et al*., 2000, 2003), the multiple start simplex (MSX) and local simplex methods (Gan and Biftu, 1996), and simulated annealing (Thyer *et al*., 1999). The shuffled complex evolution method (SCE-UA) (Duan *et al*., 1992) is widely used for hydrologic applications, and is also used in this study to minimize objective functions for calibration in SWAT.

The objective function, or difference measure between simulated and observed values, is strongly determinative for the eventual calibration result. Most hydrological studies deal with objective functions (OFs) and calibration in the time domain. Spectral calibration is less common, but there is a renewed interest in this topic (Islam and Sivakumar 2002; Fleming *et al*., 2002; Montanari and Toth, 2007). Under the assumption of a time-invariant system, this technique allows to quantify system parameters through comparison of the spectral representation (or autocorrelation properties) of simulations and measurements, even if the latter are only available in old or sparce data records or in other, but similar, regions than the study area. This is beneficial for predictions in ungauged basins (Montanari and Toth, 2007).

It has been demonstrated (Boyle *et al*., 2000; Houser *et al*., 2001) that different objective functions lead to different optimal model parameter values. Because the exact probabilistic model and observation error properties are often unknown and hence not exploited in the objective functions, it is difficult to find statistically optimal parameter sets (Gupta *et al*., 1998). The multi-objective parameter set equivalence is sometimes referred to as 'pareto-optimal' or 'non-dominant'. Currently, there is a tendency to use various objective functions simultaneously, to constrain the optimal parameter values and hence the predictive uncertainty (Yapo *et al*., 1998; Madsen, 2000; Madsen *et al*., 2002; Vrugt *et al*., 2003; Ajami *et al*., 2004; De Lannoy *et al*., 2006; Vrugt and Robinson, 2007). Besides the variety of mathematical expressions

* Correspondence to: Gabriëlle J. M. De Lannoy, Laboratory of Hydrology and Water Management, Ghent University, Coupure links 653, B-9000 Ghent, Belgium. E-mail: Gabrielle.DeLannoy@UGent.be

for the difference measures, there are generally various variables for which the objective function can be constructed, such as soil moisture and soil temperature, or fluxes such as discharge or evapotranspiration. Unfortunately, the traditional calibration at the outlet of a catchment cannot guarantee that internal processes in the system are simulated correctly (Refsgaard, 1997; Anderton *et al*., 2002). Incorporation of information from additional variables always provides an added value to the calibration. Furthermore, the selected calibration period and resolution (in space and time) of the observed and simulated information is of importance (Gan *et al*., 1997; Brath *et al*., 2004; Zehe and Blöschl, 2004). Finally, the resulting parameter values will depend on the other parameter values (multi-parameter interaction), the model structure, the input data accuracy, and the initial conditions. For the latter, it is advised to have at least a balanced state value before calibration. In practice, models have often been started from default, measured or guessed initial conditions until a stable situation was found and the initial state was balanced (Wood *et al*., 1998). This is referred to as model spin-up or warming-up. The initial state can also be found through state estimation techniques.

The objective of this study is to compare frequency and time expressions for objective functions for the SWAT model sensitivity analysis (SA) and calibration of the Tijeral catchment in Central Chile. Three alternative objective functions were developed in the frequency domain and studied in addition to the available objective function in the time domain. All analyses use *in situ* observed daily discharge values. A site description is given in Section 2. The different objective functions are introduced in Section 5 that also describes the SA and the calibration method. The results are discussed in Section 6 as well as the presumable advantages of spectral calibration.

SITE DESCRIPTION

The Tijeral catchment is a part of the Biobío catchment. With a 24 371 km^2 area, the Biobío catchment is the third largest catchment in Chile and covers about 3% of the national continental surface area. The Biobío catchment is located in the central part of Chile near Concepción. The Tijeral catchment covers the southern part of the Biobío catchment, has an area of 2311 km^2 and is a large sub-catchment (grey area in Figure 2) of the Vergara catchment (whole area in Figure 2), which drains directly into the Biobío river. The most northern Tijeral catchment border is situated roughly 120 km to the south of Concepcíon and the most western border about 50 km from the Pacific Ocean. The Tijeral catchment does not border the coastline anywhere, but it contains a coastal mountain range at the western part, with maximum heights of 1000 m. The eastern part of the catchment is situated in the Andes Mountains with areas covered in temporary snow. The central part of the catchment (the Central Valley) is nearly flat with a mean height of approximately 200 m a.m.s.l. The highest point of the study area is situated in the Andes, and has an altitude of 1777 m. The lowest altitude (68 m) is found at the outlet of the Tijeral river. Figure 1 depicts a digital elevation model (DEM) of the Tijeral basin, viewing southwards.

DATA

This study uses daily data of precipitation and discharge from January 1, 1992 through December 31, 2002 (11 years or 4018 days). Figure 2 locates eight raingauges in (and near) the Tijeral catchment with valid daily data available for the selected period, i.e. Angol, Collipulli, Encimar, Ercilla, Laguna, Lumaco, Traiguen and Tranaman. A linear relationship between the time-mean (over the selected study period) daily precipitation <P> [mm] and the altitude a [m a.m.s.l.] of these raingauges was found ($<P> = 0{\cdot}0061a + 2{\cdot}31$) with a determination coefficient (R^2) of 0·97. No significant correlation was found between the number of rainy days and the altitude ($R^2 = 0{\cdot}05$). Higher altitudes thus signify more intensive showers.

Figure 3 shows the spatially averaged precipitation (over eight rain gauges) and discharge data of the Tijeral catchment. Annual totals of precipitation and discharge are printed as well in Figure 3. Both the precipitation and discharge time series show a clear annual pattern,

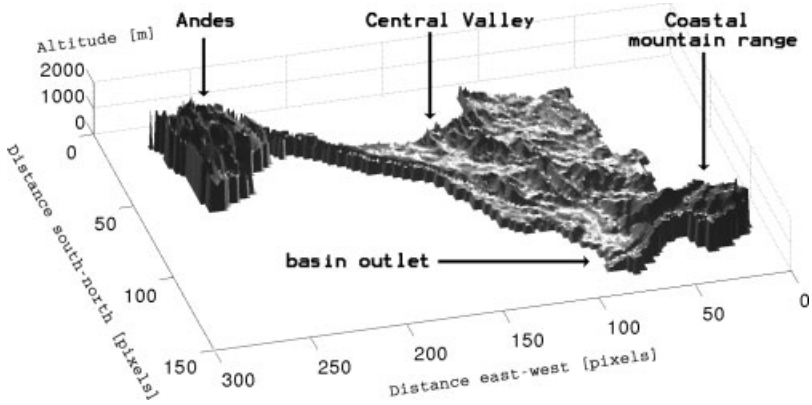

Figure 1. DEM of the Tijeral catchment, viewing southwards. The spatial resolution is 450 m

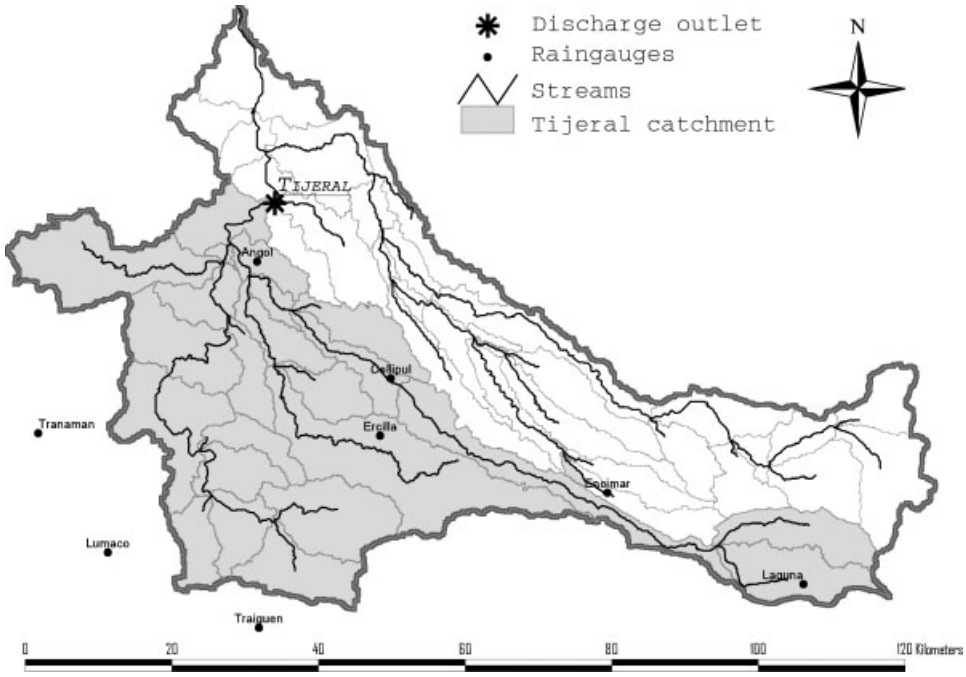

Figure 2. Location of the Tijeral catchment (grey coloured, containing 24 sub-basins) in the Vergara catchment, together with the locations of the eight used rain gauges in the neighbourhood of the Tijeral catchment

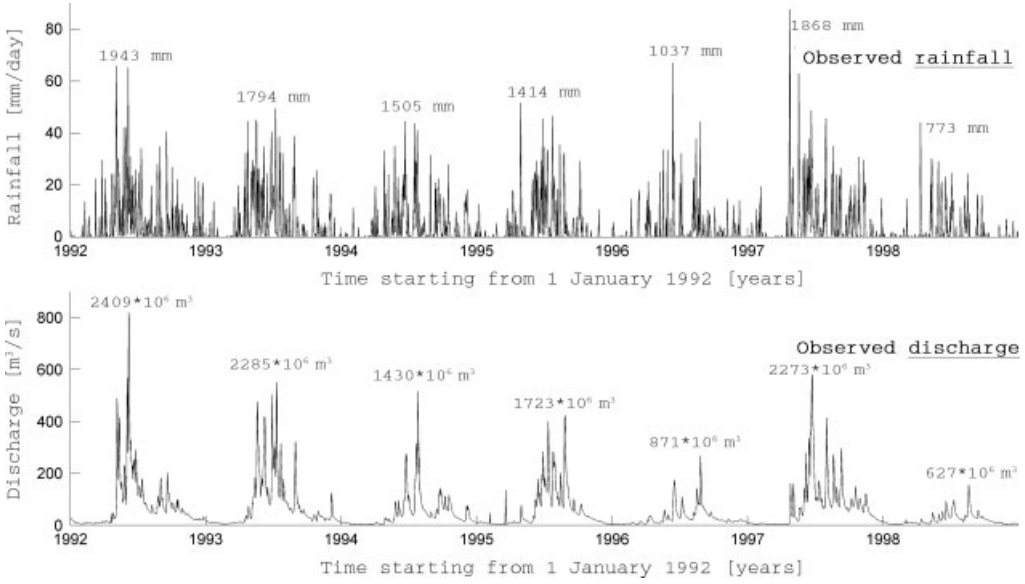

Figure 3. Time series of spatial mean rainfall and discharge time series at a daily time step for the period January 1, 1992 through December 31, 1998, collected in the Tijeral catchment from eight rain gauges and at the Tijeral outlet, respectively. Annual totals of rainfall and discharge are indicated as well

which is common in Mediterranean climate systems like the Tijeral basin. Because this catchment is situated south of the equator, and subject to a mediterranean climate, precipitation and discharge are minimal around January (summer). The years 1996 and 1998 show a distinct decline in precipitation and discharge, possibly resulting from El Niño events that occurred during those two years (Mason *et al*., 1999). The runoff ratio of the Tijeral catchment is 0·49, which is a higher value than the mean South American runoff ratio of 0·38 (Hornberger *et al*., 1998). The daily minimum and maximum air temperatures needed to force the SWAT model were obtained from three temperature gauges located in and around the Tijeral catchment. Other input data, like a DEM with a resolution of 90 m × 90 m and a soil type and land use map of the Tijeral catchment as vector images are also available. The DEM was produced during the Shuttle Radar Topography Mission (SRTM) in February, 2000. The land use map dates from 1997.

MODEL DESCRIPTION

The SWAT (Arnold *et al*., 1998), is a river basin model originally developed for the United States Department

of Agriculture (USDA) Agricultural Research Service (ARS) and Texas A&M University. A spatially semi-distributed precipitation-runoff model, such as SWAT, divides a catchment into smaller discrete calculation units for which the spatial physical property variation is limited, and hydrological processes can be treated as being homogeneous. SWAT performs two major types of catchment breakdowns into smaller units. Firstly, SWAT delineates a sub-catchment (or sub-basin) for each river network branch (Figure 2). The water routing occurs at this catchment sub-division level. Starting with the sub-basins containing the river branches of the lowest order (order 1), the discharge amount simulated at each sub-basin outlet is added to the simulated discharge of the downstream connected sub-basin, until the major catchment outlet is reached, resulting in the total catchment discharge. The next catchment breakdown appears by overlaying a soil type and land use map within sub-basin boundaries, leading to a number of unique land use and soil type combinations. Each combination is considered as a homogeneous physical property, i.e. a hydrological response unit (HRU) on which surface runoff and base flow are calculated. Surface runoff and base flow amounts of all HRUs within a sub-basin are added together, contributing to the sub-basin discharge. SWAT is thus spatially semi-distributed at the sub-basin level, but not at the HRU level. The only spatially related HRU property is that it is located within a certain sub-basin. Inside a particular sub-basin, spatially disconnected HRUs of the same type will be defined as a single entity, but they will be defined as different entities if they are situated in different sub-basins. There are 143 HRU entities in the Tijeral SWAT project. The HRU water balances are computed on a daily time step. An in-depth description of the SWAT model is given by Neitsch *et al*. (2002) and Arnold *et al*. (1998).

PARAMETER OPTIMIZATION

Introduction

Equation (1) explains the principle of discrete Fourier analysis (Thibos, 2003). A discrete time series of discharge $Q(t)$, $\forall\ t \in [1, \cdots, D]$, with D time steps, can be decomposed into a number of sine and cosine waves as follows:

$$Q(t) = \sum_{k=0}^{N} \Psi(k) \left[a(k) \cos\left(\frac{2\pi k}{D}(t-1)\right) + b(k) \sin\left(\frac{2\pi k}{D}(t-1)\right)\right] \quad (1)$$

$$= \sum_{k=0}^{N} \Psi(k) \left[c(k) \cos\left(\frac{2\pi k}{D}(t-1) - \phi(k)\right)\right] \quad (2)$$

with $N+1$ Fourier coefficients a and b $\forall\ k \in [0, N]$. The coefficients c and ϕ can be written as,$\forall\ k \in [0, N]$:

$$c(k) = \sqrt{a^2(k) + b^2(k)} \quad (3)$$

$$\phi(k) = \arctan\left(\frac{b(k)}{a(k)}\right) \quad (4)$$

where $\Psi(k) \in \left[\frac{1}{2}, 1, 1, \ldots, 1, 1\right]$ if D is an odd number and where $\Psi(k) \in \left[\frac{1}{2}, 1, 1, \ldots, 1, \frac{1}{2}\right]$ if D is an even number. For each day, $Q(t)$ can be seen as a sum of sampled trigonometric functions. As can be seen in Equation (1), there are $2 \times (N+1)$ different trigonometric functions which build the discharge series in time. The function for $k = 0$ represents a special case: $\frac{1}{2}[a(0)\cos(0) + b(0)\sin(0)] = \frac{1}{2}a(0) = m$, where m is the time series mean value. N is the number of frequencies (k is the frequency index) of the trigonometric functions. The waves corresponding to these N frequencies can also be expressed as N harmonics starting from the lowest frequency or ground harmonic ($k = 1$) which is the wave having a wavelength of exactly the time series length, and ending with the Nyquist frequency or the N^{th} harmonic ($k = N$).

According to Shannon (1949), the highest harmonic wave ($k = N$) which can build a discrete time signal, must contain at least two samples per wavelenght ($\Rightarrow N \leq \frac{D}{2}$). The N pairwise trigonometric functions (a sine and a cosine function per frequency) can be simplified to N trigonometric functions according to Equation (2), since the sum of a sine and a cosine wave of the same frequency can be described as a single cosine function. $\Psi(k)a(k)$, $\Psi(k)b(k)$ and $\Psi(k)c(k)$ stand for the amplitudes of respectively the original sine wave, the original cosine wave and the combined (cosine) wave for a given k^{th} harmonic. Equation (4) describes the phase shifts between the set of new cosine functions (Equation (2)) and the original ones (Equation (1)). The collection of amplitudes $\Psi(k)c(k)$ of Equation (2) is also known as the Fourier amplitude spectrum of a time series.

Figure 4 depicts the Fourier amplitude spectra of precipitation and discharge for $k > 0$, for the 7-year daily Tijeral basin time series shown in Figure 3. Figure 4 shows that the 7^{th} harmonic of both discharge and precipitation time series has the largest amplitude of the Fourier amplitude spectrum for $k > 0$, representing the annual (one seventh of the length of the time series) discharge (or precipitation) wave. The time series means, $\Psi(0)c(0)$, are 4·04 mm and 52·59 m^3/s for the respective precipitation and discharge data mentioned above, meaning that $\Psi(0)c(0)$ has the largest amplitude of the precipitation Fourier amplitude spectrum and that $\Psi(0)c(0)$ has the second largest amplitude of the discharge Fourier amplitude spectrum. For precipitation, the high frequency components keep a higher amplitude than for discharge, indicating more high frequency components and a larger variation in precipitation (and hence need for a higher frequent sampling to capture the entire signal) than in discharge. Figure 5 shows the Tijeral basin discharge time series, an estimation of this time series with the sum of the 0^{th} harmonic and the 7^{th} harmonic, and an estimation based on the sum of the first seven harmonics of this time series added to its time-mean value (0^{th} harmonic). This figure shows that the 7^{th} harmonic (1-year wavelength) represents the annual discharge pattern, and that the lower harmonics ($k \leq 6$) are building the rough pattern in mean

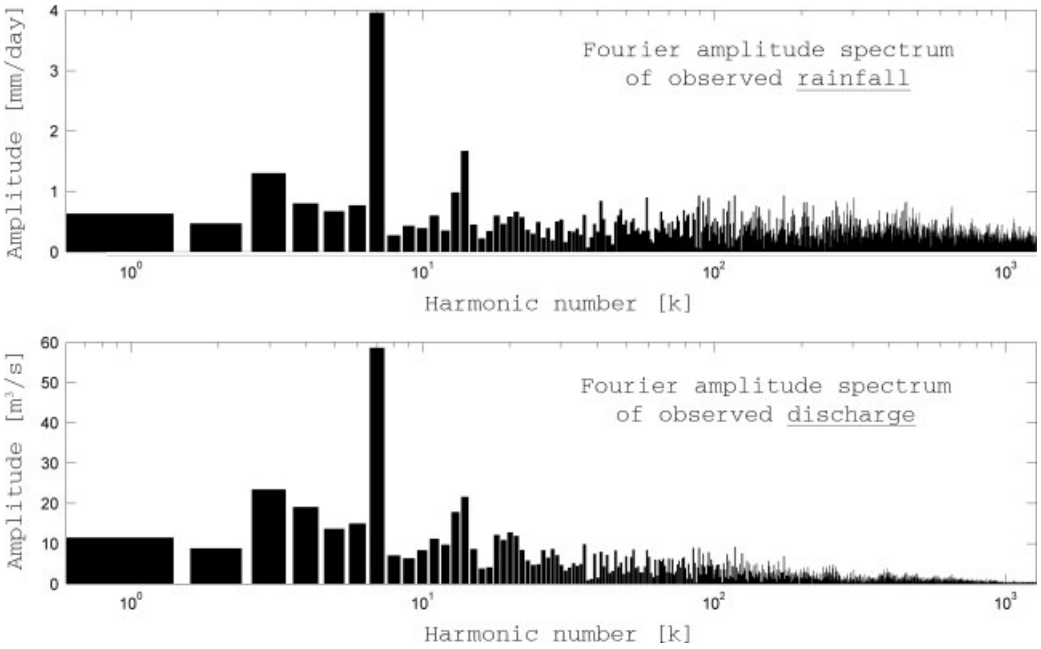

Figure 4. Fourier amplitude spectrum ($\Psi(k)c(k)$; $\forall\, k \in [1, \cdots, N]$) of (upper sub-figure) the daily spatial mean precipitation and (lower sub-figure) the discharge in the Tijeral catchment during the period January 1, 1992 through December 31, 1998. Time-mean values ($k = 0$) are not visible on the spectra because of the logarithmic nature of the abscess-axis. $\Psi(0)c(0)$ equals 4·04 mm/day for precipitation, and $\Psi(0)c(0)$ is 52·59 m^3/s for the discharge time series

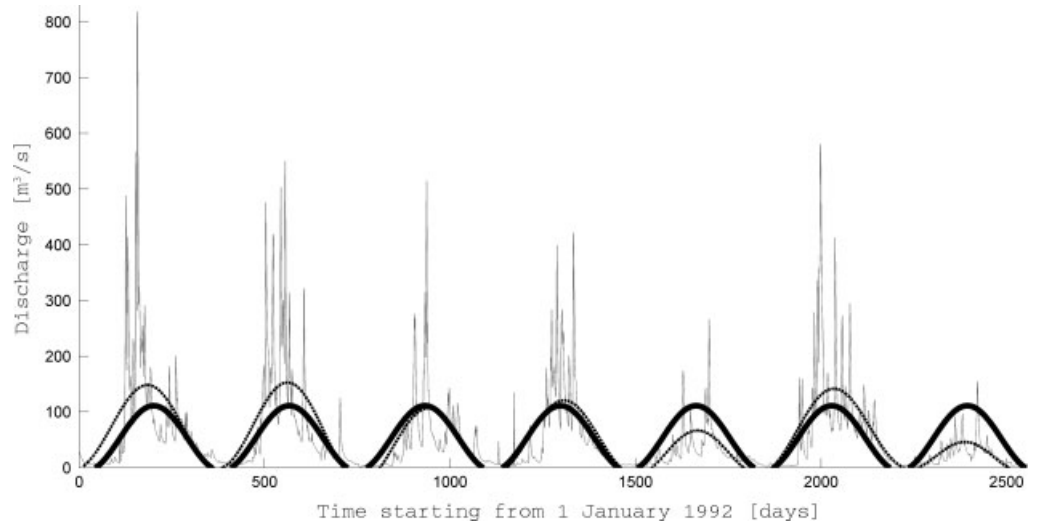

Figure 5. Discharge time series at the Tijeral outlet (grey) for period January 1, 1992 through December 31, 1998, together with a first approximation (the mean discharge plus the sine wave obtained using the 7th harmonic of the time series (thick line)) and a second approximation (the time-mean discharge plus the sum of the sine waves obtained using the seven first harmonics of the time series (dotted thin line)). Phase shifts are included

annual discharge values. The 0th harmonic is responsible for properly shifting the waves along the ordinate-axis.

Objective functions

The default objective function OF_0 used in SWAT consists of the sum of the squared residuals between the observed (Q_{obs}) and the simulated (Q_{sim}) discharge time series over a period of D days:

$$OF_0 = \sum_{t=1}^{D} [Q_{obs}(t) - Q_{sim}(t)]^2 \tag{5}$$

Here the actual observation-minus-simulation value differences over a series of time steps are minimized.

Three additional empirical spectral objective functions were developed for SWAT, more specifically OF_1, OF_2 and OF_3, which minimize spectral observation-minus-simulation component differences over a series of frequencies, as follows:

$$OF_1 = \sum_{k=0}^{N} \Psi^2(k)[c_{obs}(k) - c_{sim}(k)]^2 \tag{6}$$

$$OF_2 = \sum_{k=0}^{N} \Psi^4(k)[c^2_{obs}(k) - c^2_{sim}(k)]^2 \tag{7}$$

$$OF_3 = \sum_{k=0}^{N} \Psi^2(k)([a_{obs}(k) - a_{sim}(k)]^2 + [b_{obs}(k) - b_{sim}(k)]^2) \quad (8)$$

OF_1 is the sum of the squared differences between the observed and simulated discharge harmonic amplitudes (Equation (6)). OF_1 is thus a measure of goodness-of-fit between the observed and simulated discharge Fourier amplitude spectra. OF_2 represents the sum of the squared differences between the squared observed and simulated discharge harmonic amplitudes (Equation (7)). Since a squared Fourier amplitude spectrum defines the periodogram, which is a good approximation of the power spectral density (= Fourier transformation of the time series autocorrelation function), OF_2 is a measure of goodness-of-fit between the observed and simulated discharge spectral densities. OF_3 compares both Fourier series sine and cosine amplitudes (Equation (2)) of the observed and simulated discharge time series (Equation (8)). Since both objective functions OF_1 and OF_2 use c values, these objective functions do not account for phase shifts between the observed and simulated wave at each frequency k. Hence some information in the discharge time series is lost after Fourier transformation when used in the objective functions OF_1 and OF_2: the discharge time series cannot be built properly with the information in the Fourier amplitude spectrum or the periodogram alone. Objective function OF_3 compares all the observed and simulated discharge Fourier series components, and is thus the only tested spectral objective function which compares observed and simulated data without loss of any information present in the original discharge time series. In summary, OF_1, OF_2 and OF_3 measure the goodness-of-fit between the observed and simulated discharge series in three common spectral representations: the Fourier amplitude spectrum, the periodogram, and the Fourier series, respectively. Both the standard and newly implemented OFs for SWAT were minimized through the shuffled complex evolution algorithm (see the following text).

This paper focuses only on a selection of simple quadratic distance measures, without taking into account the exact model and observation error statistics. A wide range of alternative OFs of this type can be proposed. In literature, more formal statistical approaches to a spectral OF have been proposed, e.g. based on maximizing Whittle's Likelihood of a particular parameter set for the model (Whittle, 1953), where a statistical description of the model error is included. This approach translates into a minimization of OF_{WL} (Montanari and Toth, 2007), given by

$$OF_{WL} = \sum_{k=0}^{N} \left(\log[S_{sim}(k) + S_\Phi(k)] + \frac{S_{obs}(k)}{S_{sim}(k) + S_\Phi(k)} \right) \quad (9)$$

Here $S_{sim}(k)$ and $S_{obs}(k)$ $\forall k \in [0, N]$ are the spectral densities of the simulated and observed time series. $S_\Phi(k)$ $\forall k \in [0, N]$ is the spectral density of the autoregressive operator Φ, which takes the variance and correlation in the model residuals into account. $S_\Phi(k)$ can be calculated directly as the Fourier transform of the autocorrelation function, while for practical purposes the spectral densities for the simulations and observations are typically estimated by periodograms, i.e. $S_{sim}(k) \simeq \Psi^2(k)c^2_{sim}(k)$ and $S_{obs}(k) \simeq \Psi^2(k)c^2_{obs}(k)$. This expression 9 is based on the assumption that the model residuals are independent of the model simulation, so that the simulation-plus-residual spectral density can be computed by summing their individual spectral densities, i.e. $S_{sim}(k) + S_\Phi(k)$. In theory, the Whittle's likelihood-based OF may be more statistically correct for spectral calibration. However, it does not contain any information about phase shifts and operational model calibration requires an introduction of some assumptions about the Φ operator (both on its structure and parameters) describing the model error. Model error is *a priori* unknown and hence a calibration of the model error operator should be performed along with the calibration of the hydrological model. This issue is avoided by using the simpler OF_1, OF_2 and OF_3. However, the model error term does not need to restrict the use of OF_{WL}. One could remove the model error from Equation (9) to yield another OF, but it was chosen to limit OFs to quadratic distance measures Equations (6) through (8) only in this paper.

Sensitivity analysis

An SA quantifies the relative change in the model results due to a relative change of a model parameter. In this study, the relative change in the objective function (e.g. OF_0) value is quantified. SAs allow to select the most influential parameters on the model output or model performance. There are two major reasons to limit the calibration to the most sensitive parameters: (i) it is computationally too expensive to calibrate the model for all parameters and (ii) the calibration problem is typically underdetermined (Beven, 1989; Duan *et al.*, 1992; Yapo *et al.*, 1998; Franks *et al.*, 1999), i.e. there are mostly not enough data available to uniquely constrain all parameters. Therefore, for each OF, the most sensitive parameters will be sought to optimize the calibration with each specific OF, while leaving the less sensitive parameters uncalibrated.

In SWAT, an SA combines a one-factor-at-a-time (OAT) design and a latin-hypercube (LH) sampling by taking the latin-hypercube samples as initial points for an OAT design (van Griensven *et al.*, 2005). The OAT design quantifies the change of model output (or model performance) due to changes of only one parameter at a time, while leaving the other model parameters unaffected. This allows attributing the quantity of change in model output (or model performance) unambiguously to an individual parameter value change. Since the change in model outcome due to a parameter change generally

depends on the initial value of all model parameters, the result of one OAT analysis can only be used to calculate merely one sample of the global sensitivity (i.e. partial effect) of the model parameters. Therefore, averaging several partial effects resulting from distinct initial parameter values expresses a more global and robust form of sensitivity. SWAT uses the LH sampling method for defining distinct initial values to the model parameters which are subject to an SA. The LH sampling sub-divides the parameter range of each parameter into X intervals, where each interval has a probability of occurrence equal to $\frac{1}{X}$. If the values of a parameter set are defined as a parameter vector, then the LH sampling generates and assigns X different parameter vectors (or LH points) for the model parameter set randomly, with the restriction that each parameter has to be assigned precisely one value in each of its intervals. The LH-OAT method consists of using X LH points as initial parameter vectors for OAT analyses. In this way, the global sensitivities are calculated as the average of several (X) distinct local sensitivities (partial effects), calculated over the entire defined parameter range, resulting in a more robust way to specify parameter sensitivities. With Z model parameters, the LH-OAT analysis of SWAT thus acquires $(Z+1)\times X$ model simulations. Partial effects ($S_{i,j}$) of an SA are calculated according to:

$$S_{i,j} = \left| \frac{100\times \begin{pmatrix} \boldsymbol{M}(\mathrm{e}_1(j),\cdots,\mathrm{e}_i(j)\times(1+f_i),\cdots,\mathrm{e}_Z(j)) \\ -\boldsymbol{M}(\mathrm{e}_1(j),\cdots,\mathrm{e}_i(j),\cdots,\mathrm{e}_Z(j)) \\ \frac{1}{2}[\boldsymbol{M}(\mathrm{e}_1(j),\cdots,\mathrm{e}_i(j)\times(1+f_i),\cdots,\mathrm{e}_Z(j)) \\ +\boldsymbol{M}(\mathrm{e}_1(j),\cdots,\mathrm{e}_i(j),\cdots,\mathrm{e}_Z(j))] \end{pmatrix}}{f_i} \right| \quad (10)$$

$\forall\ i \in [1,\cdots,Z]$ and $\forall\ j \in [1,\cdots,X]$, where $\boldsymbol{M}()$ refers to the performance of a SWAT simulation (value of OF), f_i is the fraction by which the parameter e_i is changed (a predefined constant) and j in $\mathrm{e}_i(j)$ indicates that the current value of parameter e_i belongs to interval j. The sensitivity indices S_i are calculated by $S_i = \sum_{j=1}^{X} S_{i,j}/X$. For this study, the SA was setup with $f_i = 0{\cdot}05$, $X = 10$ and $Z = 27$. Owing to the excessive number of SWAT parameters, only the 27 parameters (and their default ranges) for discharge simulation are used in the SA & calibration tool. A more in depth description can be found in the work of van Griensven (2005) and in van Griensven *et al.* (2005).

In this paper, four SAs were performed, using the objective functions OF_0, OF_1, OF_2 and OF_3. For each objective function, the resulting 15 most sensitive parameters and corresponding sensitivity index (S) values are given in Table I. The lower the rank (column 1) the lower the S of this parameter. The mentioned parameters are described in Table II. Table I shows that the SAs using OF_2 and OF_3 select the same six most sensitive parameters, which are also ranked identically. SAs using OF_0 and OF_1, on the other hand, each produce a different parameter selection and ranking of the six most sensitive parameters. However, three parameters are selected by all SAs as belonging to the six most sensitive ones: CN_{II}, alpha_bf and sol_awc. CN_{II} is the most sensitive parameter for all the SAs using a spectral objective function, and the second most sensitive parameter for the SA using the default time domain objective function OF_0. CN_{II} is defined as the initial curve number for antecedent moisture condition II (AMC_{II}). Curve numbers were introduced by the USDA Soil Conservation Service to forecast surface runoff quantities $Q_X(t)$ [mm H_2O] for storm events at time t (Hjelmfelt, 1991), using precipitation P(t) [mm H_2O] as the sole

Table I. Parameters ranked according to decreasing sensitivity index S using objective functions OF_0, OF_1, OF_2 and OF_3

Rank	OF_0		OF_1		OF_2		OF_3	
	Parameter	S	Parameter	S	Parameter	S	Parameter	S
1	ch_k2	10·7	CN_{II}	7·23	CN_{II}	3·62	CN_{II}	2·66
2	CN_{II}	6·00	alpha_bf	2·09	surlag	2·67	surlag	2·41
3	smtmp	2·96	gwqmn	0·916	alpha_bf	2·42	alpha_bf	0·858
4	smfmx	2·95	sol_awc	0·913	sol_awc	0·619	sol_awc	0·444
5	alpha_bf	2·51	rchrg_dp	0·910	ch_k2	0·572	ch_k2	0·332
6	sol_awc	2·14	smfmn	0·852	gwqmn	0·490	gwqmn	0·313
7	sol_z	0·914	surlag	0·721	rchrg_dp	0·419	sftmp	0·279
8	ch_n	0·723	ch_k2	0·596	sol_k	0·227	smfmn	0·242
9	canmx	0·543	sftmp	0·422	sftmp	0·226	ch_n	0·214
10	surlag	0·383	gw_delay	0·261	smfmn	0·202	rchrg_dp	0·199
11	sol_k	0·238	canmx	0·212	ch_n	0·162	sol_k	0·135
12	slope	0·137	ESCO	0·187	gw_delay	0·135	smtmp	0·120
13	gwqmin	0·123	sol_k	0·183	ESCO	0·104	gw_delay	0·0843
14	sftmp	0·120	sol_z	0·165	slope	0·0978	sol_z	0·0741
15	rchrg_dp	0·079	slope	0·157	smtmp	0·0955	ESCO	0·073

Table II. SWAT parameters involved in sensitivity and calibration analyses

Parameter	Description	Unit	Level
alpha_bf	Base flow alpha factor	[days]	hru
ch_k2	Channel bed effective hydraulic conductivity	[mm/h]	sub
ch_n2	Manning's n value for main channel	[−]	sub
CN_{II}	Initial SCS CN_{II} value	[−]	hru
ESCO	Soil evaporation compensation factor	[−]	bsn
gw_delay	Groundwater delay	[days]	hru
gwqmin	Threshold water depth in the shallow aquifer for flow	[mm]	hru
rchrge_dp	Deep aquifer percolation fraction	[−]	hru
sftmp	Snowfall temperature	[°C]	bsn
slope	Average slope steepness	[m/m]	hru
smfmn	Melt factor for snow on December 21	[mm water/°C × day]	bsn
smfmx	Melt factor for snow on June 21	[mm water/°C × day]	bsn
smtmp	Snow melt base temperature	[°C]	bsn
sol_awc	Available water capacity	[mm water/mm soil]	hru
sol_k	Saturated soil conductivity	[mm/h]	hru
sol_z	Soil depth	[mm]	hru
surlag	Surface runoff lag coefficient	[−]	bsn

Level stands for the level at which the parameter operates: basin (bsn), sub-basin (sub) or HRU (hru).

input parameter:

$$Q_X(t) = \frac{\left[\mathrm{P}(t) - 0{\cdot}2\left(\frac{1000}{\mathrm{CN}_X} - 10\right)\right]^2}{\mathrm{P}(t) + 0{\cdot}8\left(\frac{1000}{\mathrm{CN}_X} - 10\right)}$$
$$\forall\ \mathrm{CN}_X\ \in [\mathrm{CN_I}, \mathrm{CN_{II}}, \mathrm{CN_{III}}] \qquad (11)$$

Since the curve number attributed to a land area is dependent on the area properties, curve numbers in SWAT are assigned in a spatially semi-distributed way. The major factors determining the curve number involve hydrologic soil group, cover type, treatment, hydrologic condition and antecedent moisture condition (AMC). These land properties are all medium to long-term properties, and therefore the curve number is a stable attributed variable, except for AMC. The AMC can be sub-divided into three states: dry soil (AMC_I), average soil moisture (AMC_{II}) and wet soil (AMC_{III}), based on the five-day antecedent rainfall (McCuen, 1982). Therefore, a land area homogeneous in runoff properties has three curve numbers associated with it: CN_I, CN_{II} and CN_{III}, linked with respectively AMC_I, AMC_{II} and AMC_{III}. However, CN_I and CN_{III} can both be calculated through multiplication of CN_{II} with a known factor, leaving CN_{II} as the only curve number that has to be assigned. Because CN_{II} is responsible for surface runoff simulation, it is straightforward that the sensitivity is high for this parameter. Alpha_bf is an index of groundwater flow response to changes in recharge of the discharge contributing aquifer (Smedema and Rycroft, 1983). Higher alpha_bf values signify higher groundwater flow rates. Sol_awc is the available soil water capacity.

Calibration

SWAT optimizes parameters through OF minimization with the SCE-UA method. A detailed description of this method can be found by Duan *et al.* (1994). The method can be considered a genetic algorithm, and is based on the evolution of a predefined number of parameter combinations (s) towards the global optimum. These s parameter combinations, referred to as points, are ranked in increasing order of the objective function to be minimized, and subdivided into a number of complexes (p). The first complex gets the best point, the second the second best, and so on until complex p. The first complex then gets point $p+1$, and so on, until all points have been assigned a complex. Each complex then evolves following the competitive complex evolution algorithm. After this, the obtained points are re-assigned a complex following the procedure explained above, and the algorithm is repeated until either convergence has been achieved or a predefined number of iterations has been reached. All optimizations were performed using the same random initial starting point.

The spatially semi-distributed SWAT model has both parameters having spatially distributed values and general study area parameters (see 'level' in Table II). The spatially semi-distributed parameters can be sub-divided into two categories: sub-basin parameter can be assigned a different value for each sub-basin, and HRU parameters can have a different value for each HRU. The third category of parameters in SWAT is non-spatial (e.g. basin parameters), i.e. they have only one uniform value for the entire catchment. The way SWAT changes parameter values during calibration depends strongly on which category the parameter belongs to. The (auto)calibration tool repetitively applies the model for different parameter values, and retains the parameter set that produces the most effective model result. Basin parameter values can be changed directly by replacing the initial parameter value by other values within a predefined parameter range. In this way, values which are assigned to a basin parameter during calibration are never linked to the initial parameter value. Basin parameters can be changed also by adding (or subtracting) a value within a predefined range to the initial

Table III. Most sensitive 'par' parameters, their 'dflt' default and 'cal' calibrated value (or correction factor) for different objective functions (OF_0, OF_1, OF_2, OF_3)

Objective function							
OF_0	par	ch_k2	CN_{II}	smtmp	smfmx	alpha_bf	sol_awc
	method	by rep	by mul	by rep	by rep	by rep	by mul
	dflt value	0·000	★	0·500	4·500	0·048	★
	cal value	0·004	1·056	2·886	3·967	1·334	1·525
OF_1	par	CN_{II}	alpha_bf	gwqmn	sol_awc	rchrg_dp	smfmn
	method	by mul	by rep	by rep	by mul	by rep	by rep
	dflt value	★	0·048	0·000	★	0·000	4·500
	cal value	0·977	2·692	0·018	1·622	0·309	4·949
OF_2	par	CN_{II}	surlag	alpha_bf	sol_awc	ch_k2	gwqmn
	method	by mul	by rep	by rep	by mul	by rep	by rep
	dflt value	★	4·000	0·048	★	0·000	0·000
	cal value	1·165	8·965	0·294	1·387	134·960	2217·800
OF_3	par	CN_{II}	surlag	alpha_bf	sol_awc	ch_k2	gwqmn
	method	by mul	by rep	by rep	by mul	by rep	by rep
	dflt value	★	4·000	0·048	★	0·000	0·000
	cal value	0·967	1·380	2·544	1·517	0·000	1010·900

'method' describes the method by which the default parameters values are changed during calibration: by multiplication (by mul) or by replacement (by rep) of a default parameter value. ★ indicates that the components of a distributed parameter are assigned different default pre-calibrated values in space. Calibrated distributed parameters are obtained by multiplying the default parameter component values with the factor which is given after 'cal value' for spatial distributed parameters.

parameter. In this method, initial parameters are reference points to which a value is added or subtracted during each calibration step. To change component values of spatially semi-distributed parameters (sub-basin and HRU parameters), it is common to multiply all component values of such a parameter by a common factor (The label 'by mul' in Table III indicates that the parameter is changed by multiplication). In this way the component values of the spatially semi-distributed parameters retain their spatial properties during the calibration. A disadvantage of this method is that the proportions between the different component values of the same spatial parameter are unaltered. It is possible to treat each component of a spatial parameter as a unique parameter in the calibration of SWAT, but this is not advisable because calibration time would then increase enormously. Some spatially semi-distributed parameters in SWAT are assigned by default the same uniform value. Those spatial parameters can also be changed, like basin parameters, by replacement or by addition/subtraction. The label 'by rep' in Table III indicates that the parameter is changed by replacement. The addition/subtraction method is not used in this study. Besides the most effective values found through calibration (in Table III: cal value), Table III also mentions the default values (in Table III: dflt values) of the basin parameters before calibration.

Ideally, comparison of the performances of different OFs should be performed by involving all model-parameters in the calibrations. Owing to computational restrictions only the six most sensitive parameters were calibrated for each OF, because they have the largest impact on the model outcome. In this way, different parameter sets are calibrated in different calibration experiments with different OFs, to allow the most optimal performance with each OF. Selecting a fixed parameter set for all calibration experiments would not allow for any more fair comparison. The performance of each OF would simply depend on how sensitive it is to the particular chosen parameter set and none of the results would approach the best possible outcome.

For this study, 2 years were used for warm-up, i.e. from January 1, 1992 through December 31, 1993. After this warm-up period, the actual daily calibration period was from January 1, 1994, through December 31, 1995 (2 years), which sets the lowest observable oscillation frequency and the frequency resolution to 1/730 days^{-1}. In the time domain, the sampling frequency was fixed at 1/day, which sets the highest detectable (Nyquist) frequency (365*th* harmonic) to 365/730 days^{-1}.

Table III shows that the very sensitive parameter CN_{II} is practically left unchanged relative to its initial guess through all four calibrations, meaning that the default value is near-optimal. Further, sol_awc is multiplied by approximately 1·5 for every calibration. The most effective value assignment of alpha_bf, gwqmn and ch_k2 strongly depends on the objective function used in the calibration. One can conclusively state that, except for the spatially semi-distributed parameters CN_{II} and sol_awc, the best parameter value strongly depends on the OF. This complies with the conclusions of Refsgaard (1997) and Anderton *et al.* (2002) that calibration at the outlet of a catchment cannot guarantee that internal system processes are simulated correctly, and that several model representations (different parameters) may yield reasonable discharge simulations (equifinality).

MODEL VALIDATION

Criteria

Because no single validation criterion can quantify all features of simulations, a set of four criteria was used for

validation:

$$NS = 1 - \frac{\sum_{t=wu+1}^{D} [Q_{obs}(t) - Q_{sim}(t)]^2}{\sum_{t=wu+1}^{D} [Q_{obs}(t) - < Q_{obs} >]^2} \quad (12)$$

$$RMSE = \sqrt{\frac{1}{D - wu} \sum_{t=wu+1}^{D} [Q_{obs}(t) - Q_{sim}(t)]^2} \quad (13)$$

$$R = \frac{\sum_{t=wu+1}^{D} [Q_{obs}(t) - < Q_{obs} >][Q_{sim}(t) - < Q_{sim} >]}{\sqrt{\sum_{t=wu+1}^{D} [Q_{obs}(t) - < Q_{obs} >]^2 \sum_{t=wu+1}^{D} [Q_{sim}(t) - < Q_{sim} >]^2}} \quad (14)$$

$$RB = 100 \frac{< Q_{obs} > - < Q_{sim} >}{< Q_{obs} >} \quad (15)$$

With D the total number of simulated days, wu the number of warm-up days, $Q_{obs}(t)$ and $Q_{sim}(t)$ (m^3s^{-1}) the observed and simulated discharge at day $t(t \in [1, \cdots, D])$ and $<Q_{obs}>$ and $<Q_{sim}>$ the mean observed and mean simulated discharge, $\frac{1}{D - wu} \sum_{t=wu+1}^{D} Q_{obs}(t)$ and $\frac{1}{D - wu} \sum_{t=wu+1}^{D} Q_{sim}(t)$. *NS* is the Nash-Sutcliffe (Nash and Sutcliffe, 1970) coefficient, *RMSE* represents the root mean square error, *R* is the correlation coefficient and relative bias (*RB*) expresses the difference between the mean daily observed and simulated discharge relative to the mean daily observed discharge. *RB* is solely a measure of bias. *RMSE* and *NS* are combined measures of bias and dynamics. However, *RMSE* and *NS* focus more on adjustment of dynamics than on bias minimization. *R*, finally, has no inherent bias measurement.

SWAT allows a user to choose between two methods of water routing over sub-basins: the variable storage method (default) and the Muskingum method. Because of largely degraded simulation results with default parameter values of the variable storage method ($RMSE = 82{\cdot}2$ m^3/s) in comparison with the Muskingum method ($RMSE = 32{\cdot}0$ m^3/s), the latter routing technique has been used for the remainder of this study.

Results

The validation was performed from January 1, 1998 through December 31, 2002 ($D - wu = 1826$ days). Figures 6 and 7 allow a graphical interpretation of the simulations with calibrated parameters (Table III) for the two latest years of the validation period. All OFs effectively improve the discharge simulations compared to the uncalibrated one. For example, the decreasing limb in the hydrograph from September 2001 is clearly improved. The general dynamics for the different OFs are very similar, indicating that both spectral and time domain calibration are valid options, even though OF1 and OF2 contain a weaker observation constraint (phase shift information is lost). OF_2 needs some additional attention, because it shows distinct dynamics. During periods of high discharge, the OF_2 curve shows a more intense response, while for low discharge periods small discharge peaks are smoothed out strongly and do not reflect the actual observed small peaks. During periods of low observed discharge, all simulations, except the OF_2, underestimate observed discharge rates by simulating almost zero discharge. The aforementioned simulation pattern of OF_2 can be explained by the calibrated parameter values. With OF_2, the responsible parameters for the amount and speed of precipitation runoff receive high (CN_{II}) to very high (surlag) values, enhancing a rapid response in the simulation. On the other hand, ch_k2 and gwqmn are very high for OF_2 as compared to their values with other OF calibrations or to the default values. As ch_k2 stands for transmission losses through the streambed, a large amount of stream water will contribute to the sub-stream base flow. Because transmission losses are modelled as having a constant rate [mm/h], independent on the stream depth, they will be relatively more important at lower stream depths (summer) than at higher stream depths (winter). Owing to the large modelled transmission losses of OF_2, the shallow aquifer will be rapidly filled up and will contribute a long time to sub-stream base flow, even when the stream is almost dry, resulting in a smoother but non-zero simulated curve at low observed discharge values in the summer.

OF_3 contains essentially the same information as OF_0, and more information than either OF_1 or OF_2. While OF_1 and OF_2 use the same (reduced, i.e. no phase shift) information for calibration, they produce very distinct discharge curves in the validation period. As can be checked with Equations (6) and (7), the same observation-minus-simulation amplitude difference of a certain harmonic produces a much higher OF_2 value when the observed amplitude is large than when it is small. An example: Let $c_{obs} = 4$ m^3s^{-1} and $c_{sim} = 3$ m^3s^{-1}, it then follows that $(c_{obs} - c_{sim}) = 1$ m^3s^{-1} and $(c_{obs}^2 - c_{sim}^2) = 49$ $(m^3s^{-1})^2$. If however $c_{obs} = 80$ m^3s^{-1} and $c_{sim} = 79$ m^3s^{-1}, than $(c_{obs} - c_{sim}) = 1$ m^3s^{-1} again, but $(c_{obs}^2 - c_{sim}^2) = 25\,281$ $(m^3s^{-1})^2$. This means that OF_2 emphasizes the larger amplitudes of the Fourier amplitude spectrum of observed discharge during the calibration process. The squaring of the observed-minus-simulated amplitude differences in all spectral OFs in this study puts more weight on the largest spectral errors in general (regardless of the amplitudes of the harmonics of the observed Fourier amplitude spectrum involved).

Figure 8 depicts the observed discharge Fourier amplitude spectrum for the validation period and its amplitude differences with the various simulated (uncalibrated and calibrated) Fourier amplitude spectra. Figure 8 shows that the lowest harmonics ($k < 20$) of the observed discharge have the largest values and that the highest harmonics

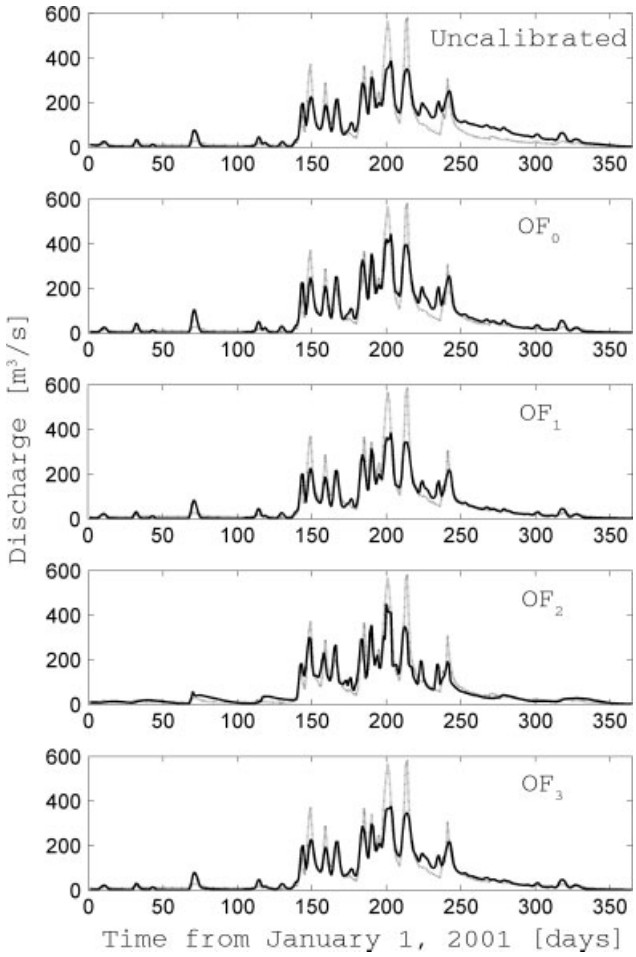

Figure 6. Observed (grey curve) and simulated (black curve) discharge time series at the Tijeral outlet for validation year 2001, for the uncalibrated, OF_0-, OF_1-, OF_2- and OF_3-calibrated simulations, respectively

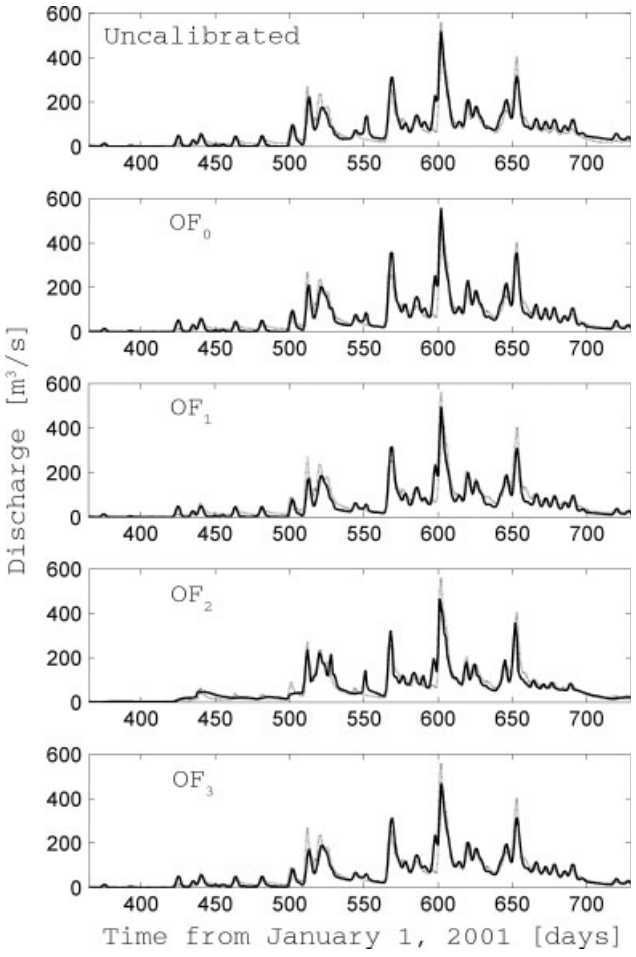

Figure 7. Observed (grey curve) and simulated (black curve) discharge time series at the Tijeral outlet for the validation year 2002, for the uncalibrated, OF_0-, OF_1-, OF_2- and OF_3-calibrated simulations, respectively

($k > 200$) are very small. Hence OF_2 changes the uncalibrated parameters in a way that will fit most of the low frequency range very well, but will largely overlook the errors on the higher frequencies. Since a change in a parameter value (e.g. during calibration) influences the whole Fourier amplitude spectrum of simulated discharge, the OF_2 simulations will produce larger errors in the higher harmonics than the other OF simulations, as can be seen in Figure 9 which shows observed-minus-simulated discharge Fourier amplitude spectra relative to the observed discharge Fourier amplitude spectrum. Figure 9 shows that the higher harmonics are simulated very differently with use of OF_2 and Figure 8 shows that the spectral errors of the high and moderate amplitude harmonics of the observed Fourier amplitude spectrum (including $k = 0$) are reduced very well for OF_2. Finally, one can notice in Figure 8 that, despite the fact that OF_2 assigns by far more weight on the error of the second harmonic (large amplitude) than OF_3 does, OF_3 is capable of reducing this error far more than OF_2, which is probably a result of including phase shifts in the calibration with OF_3.

The biased (underestimated) observed-minus-simulated spectral differences (Figures 8 and 9) at high frequency harmonics ($k > 200$) may be linked with the biased (underestimated) discharge time series simulations during periods of low discharge (see Figure 6 and Figure 7), for all simulations except for the OF_2-based simulation which produces less biased (but more random) spectral errors at the high frequency range of its Fourier amplitude spectrum and less time domain errors during periods of low discharge. In this light, the smoothing of the simulated discharge time series of OF_2 can be seen as a result of more unbiased and almost uncalibrated high frequency harmonics.

Table IV summarizes the model performance (*NS*, *RMSE*, *RB* and *R*) during the calibration and validation period, both with default and calibrated parameters resulting from the minimization of the four different objective functions for their respective most sensitive parameters. Obviously, all simulations with calibrated parameters improve model results compared to simulations with uncalibrated parameters. It is noteworthy that during the calibration and validation periods, minimization of OF_0 (squared time residuals) does not lead to the lowest *RMSE* (also squared time residuals). Leaving the values of the *RB* aside for now, Table IV shows that OF_0 and OF_1 simulations have an almost identical performance for the validation period. However, for the

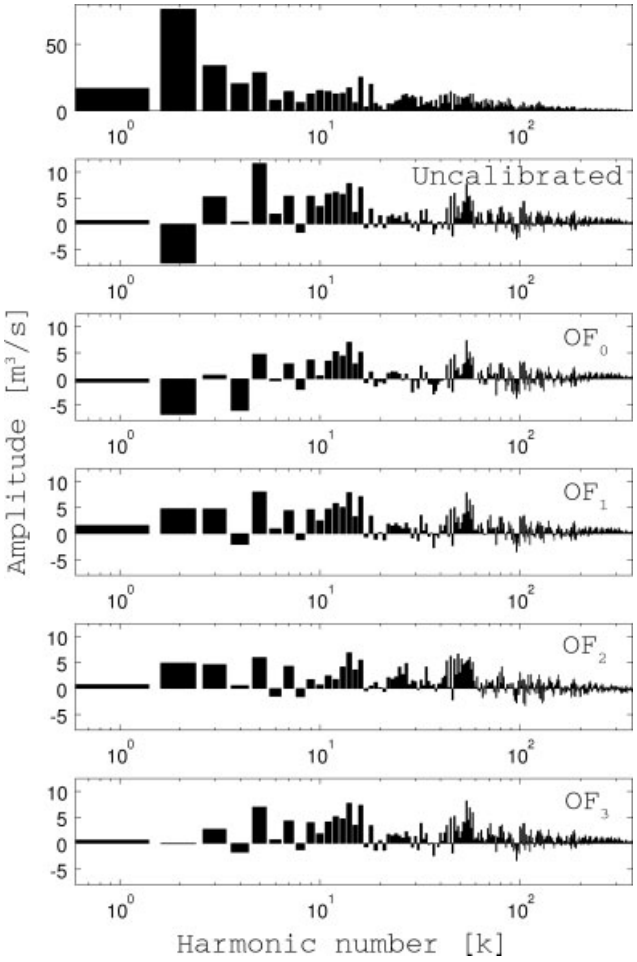

Figure 8. (Upper sub-figure) observed discharge Fourier amplitude spectrum and (other sub-figures) observed-minus-simulated (either uncalibrated or calibrated with OF_0, OF_1, OF_2, OF_3) discharge Fourier amplitude spectrum difference ($\varepsilon(k) = \Psi(k)c_{obs}(k) - \Psi(k)c_{sim}(k)$, $\forall\ k$) at the Tijeral outlet during two validation years (2001 and 2002). $\Psi(0)c_{obs}(0) = 65{\cdot}51$ m^3/s and $\varepsilon(0) = -5{\cdot}36$, $-1{\cdot}33$, $7{\cdot}77$, $0{\cdot}80$ and $2{\cdot}69$ m^3/s for the uncalibrated, OF_0-, OF_1-, OF_2- and OF_3-calibrated simulations, respectively

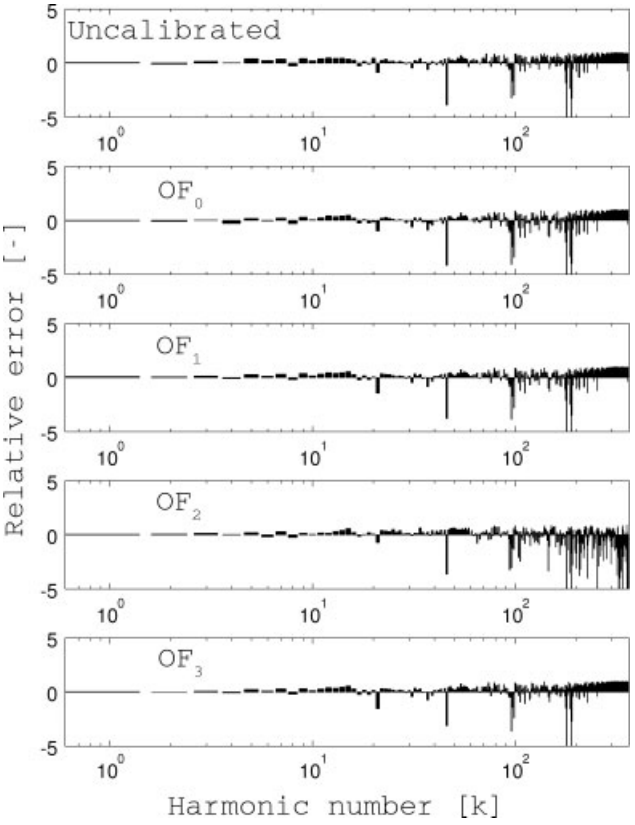

Figure 9. Relative observed-minus-simulated (either uncalibrated or calibrated with OF_0, OF_1, OF_2, OF_3) discharge Fourier amplitude spectrum difference ($\varepsilon_r(k) = (\Psi(k)c_{obs}(k) - \Psi(k)c_{sim}(k))/\Psi(k)c_{obs}(k)$, $\forall\ k$) at the Tijeral outlet during two validation years (2001 and 2002). $\varepsilon_r(0) = -0{\cdot}082$, $-0{\cdot}020$, $0{\cdot}118$, $0{\cdot}012$ and $0{\cdot}041$ for the uncalibrated, OF_0-, OF_1-, OF_2- and OF_3-calibrated simulations, respectively

calibration period, OF_1 produces better *NS* and *RMSE* results and a slightly worse *R* as compared to OF_0. A simulation with parameters calibrated by OF_2 produces the worst results on all three criteria for both the calibration and validation period, while a simulation with calibrated parameters using OF_3 produces best results on all three criteria for both simulation periods and performs substantially better than OF_0 during calibration. As mentioned in Section 5.2, OF_3 is the only OF in the frequency domain containing the complete information included in the time series of discharge. It can thus be expected that OF_3 produces the best results out of all spectral OFs.

For the *RB* criterion, there is a high variance among the most results for the different OFs for parameter calibration, except for OF_2. OF_0 restricts the bias implicitly in each term during the calibration process, but nevertheless, the *RB* is not very well restricted with use of OF_0 during the calibration period ($-10{\cdot}3\%$). *RB* is proportional to the time-mean bias, i.e. the difference between mean observed and mean simulated discharge. In OF_1 and OF_2, there thus exists only one out of $N + 1$ terms which contributes directly to the restriction of the bias, i.e. if $k = 0$ (mean value; see Section 5). However, the mean discharge value of the observed time series has in general a high amplitude compared to the amplitudes of the other harmonics of the discharge time series (49·54 m^3/s

Table IV. Error criteria values of discharge simulations using default uncalibrated values for all parameters (default), and using values of the calibrated six most sensitive parameters for different OFs, over the calibration period (January 1, 1994 through December 31, 1995) and the validation period (January 1, 1998 through December 31, 2002)

Objective function	Parameter set	*NS* [−]	*RMSE* [m^3/s]	*RB* [%]	*R* [−]
		Calibration Period			
default	default	0·800	30·0	−19·3	0·919
OF_0	sens-OF_0	0·854	25·7	−10·3	0·950
OF_1	sens-OF_1	0·887	22·5	+5·80	0·943
OF_2	sens-OF_2	0·824	28·1	−1·34	0·914
OF_3	sens-OF_3	0·902	21·0	+2·20	0·950
		Validation period			
default	default	0·814	31·8	−9·93	0·907
OF_0	sens-OF_0	0·853	28·3	−1·61	0·929
OF_1	sens-OF_1	0·853	28·3	+12·6	0·928
OF_2	sens-OF_2	0·844	29·2	−2·22	0·919
OF_3	sens-OF_3	0·865	27·1	+9·96	0·933

for the calibration period, and 65·51 m^3/s for the validation period). This means this harmonic is weighted heavily in OF_2 calibration and explains the *RB* during the calibration and the validation period. With OF_3 there exists only one out of D terms that restricts the bias in a direct manner. Although $D-1$ out of D terms in OF_3 (and N out of $N+1$ terms in OF_1 and OF_2) do not constrain the bias directly, it is also assumable that a convergence of the harmonics (except for $k=0$) of the simulated and observed discharge during calibration, indirectly causes the bias to reduce. However, it is not clear to which degree, both qualitatively and quantitatively this phenomenon takes place. One can assume that OF_1 and OF_3 are less likely to limit bias than and OF_2. Regarding the *RB* results of the calibration and the validation period, OF_2 seems to perform better than OF_0 in restricting the bias between observed and simulated discharge.

Properties of spectral calibration

The time OF_0 and spectral OF_3 result in the best performances, because they are the only OFs using the full observational constraining information during calibration. However, they produce very different results, because different discharge features are accentuated during calibration. OF_3 possibly allocates more information for simulation shape adjustment and less for bias adjustment than OF_0 does. OF_3 has an extra potential over OF_0: inside OF_3, specific weights can be given to specific Fourier coefficients (e.g. to $\frac{a(0)}{2}$ or the time-mean value) resulting in the possibility of reallocating the available discharge information for bias or shape adjustment at own preference. If one is particularly interested in either fast or long-term responses of the system to be calibrated, some spectral components can be assigned a higher or lower weight, e.g. by passing the simulated and observed discharge spectra through high-pass or low-pass filters, respectively. As explained above, high amplitude harmonics of observed discharge can also be given a significant higher weight by comparing periodograms (OF_2) instead of Fourier amplitude spectra. Although it is not investigated in this study, it may be interesting to evaluate what spectral OFs may learn about the parameter impact on model characteristics. For instance by evaluating the sensitivity of each of the spectral components to the model parameters in an SA, or by investigating whether there exist frequencies which are highly sensitive to particular parameters. The disadvantage of the spectral OFs compared to OF_0 is the need for an additional processing of both the model output and the constraining or validating observations to transform the time domain information into the frequency domain. At the other hand, if only the spectral densities are used (and not all individual Fourier coefficients) in the OF, the Fourier transform of the autocorrelation functions could be calculated instead of transforming the original time series and no complete data records are needed. It has been highlighted earlier that spectral calibration may have an interesting advantage over time domain calibration to characterize ungauged basins (Montanari and Toth, 2007), exactly because it is sufficient to have an idea of the autocorrelation function, without need for a full data time series.

Another topic of interest is the different impact of both the sampling frequency and the duration of calibration period on time and frequency calibration. A land surface system produces signals (like discharge) with a particular bandwidth. The sampling frequency determines the highest observable harmonic. In order to capture the entire signal, the sample frequency could be increased to extend the allowable signal frequency range, but there is little advantage in sampling frequencies beyond the signal bandwidth. The calibration period, at the other hand, determines the frequency resolution and the lowest detectable harmonic following the 0th one (mean value). Processes typically show different features over different time scales and it is thus important that the calibration period covers the time scale of interest. A longer calibration period increases the variability in temporal dynamics and increases the spectral resolution, but without limiting the variance in the spectral functions. Therefore, once the calibration period exceeds the most important low frequency oscillation, it might be interesting to split the time series into windows and calculate ensemble averaged spectral functions to reduce the uncertainty in the spectral representations (Welch, 1967). The duration of the calibration period and the sampling frequency will determine the weight of each spectral component in the spectral OFs. Analysis of the spectral components of a hydrological signal can help determining the optimal sample period and frequency.

CONCLUSIONS

To limit the predictive uncertainty, it is strongly recommended to optimize the tunable parameters in land surface modelling during the system identification process. Typical hydrologic calibration studies aim at minimizing the difference between observed and simulated stream discharge time series. This is exactly the default procedure for calibration of the SWAT model. In this paper, it is investigated if calibration in the frequency domain can compete with the traditional time domain calibration for precipitation-discharge processes in the Tijeral catchment in Chile.

The standard OF in SWAT measures the squared residuals between observed and simulated time series (OF_0). Three spectral OFs were proposed, based on matching Fourier amplitude spectra (OF_1), periodograms (OF_2) and individual Fourier series components (OF_3) of simulated and observed time series of discharge. The most sensitive parameter set for discharge processes was largely similar for the three different spectral OFs, but clearly distinct from the one found through an SA with the traditional time series performance of SWAT. However, three out of the six most sensitive parameters were identical for all OFs: CN_{II}, alpha_bf and sol_awc.

After a 2-year calibration period (which was proceeded by 2 years of warm-up), the best SWAT simulations (regarding the validation criterium in use: R, $RMSE$, NS or temporal mean bias) were found when using either OF_0 or OF_3 with their respective most sensitive parameters. During the calibration period, OF_3 performs equal (R) or better ($RMSE$, NS, temporal mean bias) than OF_0, but during the validation period the temporal mean bias is more constrained after calibration with OF_0. OF_3 is the spectral objective function of choice, because no temporal information is lost in the spectral representation of the discharge. The inclusion of more information in the OF will generally lead to a more constrained parameter set: any information loss will increase the uncertainty in the parameter values. Spectral calibration with OF_3 was better than time domain calibration (OF_0) during the calibration period, but both only differed marginally for simulations in the validation period, leaving both options equally worthy in this study.

Spectral calibration is slightly more complicated than time calibration, because of the need of an additional post-processing of model output. Calibration in the time domain may be the general path forward for hydrologic model optimization, but spectral calibration has some interesting advantages, which could be further explored in the future. Depending on the objective of the study, one may be interested in either fast or long-term responses of the system. To this end, one could accentuate particular spectral components in the OFs. Spectral calibration might also be interesting when only correlation properties (as opposed to actual full data records) are available for Fourier transformation to spectral densities. Finally, spectral signal analysis may help to determine the optimal sampling length and frequency to optimally constrain processes to particular features at specific time scales.

ACKNOWLEDGEMENTS

The authors wish to express their sincere gratitude to Patrick Debels and Alejandra Stehr for hosting the first author of this paper at *Centro de Ciencias Ambientalis* EULA-CHILE, Universidad de Concepción. The authors also want to thank the Chilean General Water Directorate DGA, in particular the staff and Direction of the Biobío Division; the Chilean Meteorological Directorate (DMC), the "Mininco" and "Bosques Arauco" Forestry Companies, as well as all other data providers not explicitly mentioned above. Gabriëlle De Lannoy is a postdoctoral research fellow of the Research Foundation Flanders (FWO). Alberto Montanari and three anonymous reviewers are thanked for their constructive comments.

REFERENCES

Ajami NK, Gupta H, Wagener T, Sorooshian S. 2004. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of Hydrology* **298**(1-4): 112–135.

Anderton S, Latron J, Gallart F. 2002. Sensitivity analysis and multi-response, multi-criteria evaluation of a physically based distributed model. *Hydrological Processes* **16**: 333–353.

Arnold J, Srinivasan R, Allen P. 1998. Large area hydrologic modeling and assessment. Part I: Model development. *Journal of the American Water Resources Association* **34**(1): 73–89.

Beven K. 1989. Changing ideas in hydrology: the case of physically-based models. *Journal of Hydrology* **105**: 157–172.

Beven K. 1995. Linking parameters across scales: Subgrid parameterizations and scale dependent hydrological models. *Hydrological Processes* **9**: 507–525.

Boyle DP, Gupta HV, Sorooshian S. 2000. Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. *Water Resources Research* **36**(12): 3663–3674.

Brath A, Montanari A, Toth E. 2004. Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model. *Journal of Hydrology* **291**: 232–253.

De Lannoy GJM, Houser PR, Pauwels VRN, Verhoest NEC. 2006a. Assessment of model uncertainty for soil moisture through ensemble verification. *Journal of Geophysical Research* **111**(D10): D10101, 1–18. DOI:10.1029/2005JD006367.

Doherty J. 2001. *PEST-ASP Users' Manual*. Watermark Numerical Computing: Brisbane, Australia.

Duan Q, Sorooshian S, Gupta V. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* **28**(4): 1015–1031.

Duan QY, Sorooshian S, Gupta VK. 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology* **158**(3-4): 265–284.

Fleming SW, Lavenue AM, Aly AH, Adams A. 2002. Practical applications of spectral analysis to hydrologic time series. *Hydrological Processes* **16**: 565–574.

Franks S, Beven K, Gash J. 1999. Multi-objective conditioning of a simple SVAT model. *Hydrology and Earth System Sciences* **3**(4): 477–489.

Gan TY, Biftu GF. 1996. Automatic calibration of conceptual rainfall-runoff models: optimization algorithms, catchment conditions, and model structure. *Water Resources Research* **32**(12): 3513–3524.

Gan TY, Dlamini EM, Biftu GF. 1997. Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *Journal of Hydrology* **192**: 81–103.

van Griensven A. 2005. *Sensitivity, Auto-Calibration, Uncertainty and Model Evaluation in SWAT 2005*. UNESCO: Belgium.

van Griensven A, Meixner T, Grunwals S, Bishop T, Diluzio M, Srinivasan R. 2005. A global sensitivity analysis tool for the parameters of multi-variable catchment model. *Journal of Hydrology* **324**: 10–23.

Gupta HV, Sorooshian S, Yapo PO. 1998. Toward improved calibration of hydrological models: multiple and noncommensurable measures of information. *Water Resources Research* **34**(4): 751–763.

Hjelmfelt A. 1991. Investigation of curve number procedure. *Journal of Hydraulic Engineering* **117**(6): 725–737.

Hornberger G, Raffensperger J, Wiberg P, Eshleman K. 1998. *Elements of Physical Hydrology*. The John Hopkins University Press: Baltimore.

Houser PR, Gupta HV, huttleworth J, Famiglietti JS. 2001. Multi-objective calibration and sensitivity of a distributed land surface water and energy balance model. *Journal of Geophysical Research* **106**(24): 33:421–33:433.

Islam M, Sivakumar B. 2002. Characterization and prediction of runoff dynamics: a nonlinear dynamical view. *Advances in Water Resources* **25**(2): 179–190.

Madsen H. 2000. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology* **235**(3-4): 276–288.

Madsen H, Wilson G, Ammentorp HC. 2002. Comparison of different automated strategies for calibration of rainfall-runoff models. *Journal of Hydrology* **261**(1-4): 48–59.

Mason S, Goddard L, Graham N, Yulaeva E, Sun L, Arkin P. 1999. The IRI seasonal climate prediction system and the 1997/98 El Nino event. *Bulletin of the American Meteorological Society* **80**: 1853–1873.

McCuen RH. 1982. *Guide to Hydrologic Analysis using SGS Methods*. Prentice Hall: Englewood Cliffs, New Jersey..

Mertens J, Madsen H, Kristensen M, Jacques D, Feyen J. 2005. Sensitivity of soil parameters in unsaturated zone modelling and the relation between effective, laboratory and in situ estimates. *Hydrological Processes* **19**: 1611–1633.

Montanari A, Toth E. 2007. Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins? *Water Resources Research* **43**(5): W05434, DOI:10.1029/2006WR005184.

Nash J, Sutcliffe J. 1970. River flow forecasting through conceptual models, part I: a discussion of principles. *Journal of Hydrology* **108**: 282–290.

Neitsch S, Arnold J, Kiniry J, Williams J, King K. 2002. *Soil and Water Assessment Tool Theoretical Documentation, Version 2000*. Texas Water Resources Institute: College Station, Texas. Available at: http://www.brc.tamus.edu/swat/downloads/doc/swat2000theory.pdf.

Reed P, Minsker B, Goldberg D. 2000. Designing a competent simple genetic algorithm for search and optimization. *Water Resources Research* **36**(12): 3757–3761.

Reed P, Minsker BS, Goldberg DE. 2003. Simplifying multiobjective optimization: an automated design methodology for the nondominated sorted genetic algorithm-II. *Water Resources Research* **39**(7): 1196, DOI:10.1029/2002WR001483.

Refsgaard JC. 1997. Parameterization, calibration and validation of distributed hydrological models. *Journal of Hydrology* **198**: 69–97.

Shannon CE. 1949. Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers* **13**(1): 10–21.

Smedema LK, Rycroft DW. 1983. *Land Drainage: Planning and Design of Agricultural Drainage Systems*, Cornell University Press: Ithica, New York.

Thibos L. 2003. *Fourier Analysis for Beginners*, e-book, School of Optometry, Indiana University http://research.opt.indiana.edu/Library/FourierBook/notice.html.

Thyer M, Kuczera G, Bates BC. 1999. Probabilistic optimization for conceptual rainfall-runoff models: a comparison of the shuffled complex evolution and simulated annealing algorithms. *Water Resources Research* **35**(3): 767–773.

Vrugt JA, Gupta HV, Bastidas LA, Boutem W, Sorooshian S. 2003. Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research* **39**(8): SWC5:1–SWC5:19.

Vrugt JA, Robinson BA. 2007. Improved evolutionary optimization from genetically adaptive multimethod search. *Proceedings of the National Academy of Sciences of the United States of AMerica* **104**(3): 708–711.

Welch PD. 1967. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short Modified Periodograms. IEEE Transactions on Audio and Electroacoustics AU-15, 70–73.

Whittle P. 1953. Estimation and information in stationary time series. *Arkiv fü Matematik* **2**(5): 423–434.

Wood E, Lettenmaier D, Liang X, Lohmann D, Boone A, Chang S, Chen F, Dai Y, Dickinson R, Duan Q, Ek M, Gusev Y, Irannejad P, Mitchel K, Nasonova O, Noilhan J, Schaake J, Schlosser A, Shao Y, Shmakin A, Verseghy D, Warrach K, Xue Y, Yang Z, Zeng Q. 1998. The project for intercomparison of land-surface parameterization schemes (pilps) phase 2(c) red-arkansas river basin experiment: 1. experiment description and summary intercomparisons. *Global and Planetary Change* **19**(1-4): 115–135.

Yapo P, Gupta H, Sorooshian S. 1998. Multi-objective global optimization for hydrological models. *Journal of Hydrology* **204**: 83–97.

Zehe E, Blöschl G. 2004. Predictability of hydrologic response at the plot and catchment scale: role of initial conditions. *Water Resources Research* **40**: W10202, 1–21.

